ARE WILDCAT WELL OUTCOMES DEPENDENT OR INDEPENDENT?

by

Gordon M. Kaufman* and P. J. Lee**

MIT Sloan School Working Paper 3373
January 1992

*Massachusetts Institute of Technology, Cambridge, MA 02139-4307
**Institute of Sedimentary and Petroleum Geology, Geological Survey of Canada,
3303 33rd Street NW, Calgary, Canada T26 2A7

ABSTRACT

A binary logit model is adapted to the spatial point process represented by outcomes of wildcat wells as a function of drilling history. The probability of success of the $(n + 1)$ st wildcat is made dependent on this well's location and on outcomes of wildcats previously drilled within a distance d of this well. This simple model is a device for investigating patterns of dependencies of wildcat well outcomes and for projecting probabilities of drilling success at particular locations. Application to two Canadian petroleum plays show how to use it.

KEYWORDS: spatial point process, binary logit, wildcat drilling

ARE WILDCAT WELL OUTCOMES DEPENDENT OR INDEPENDENT?

G.M. Kaufman* and P.J.Lee**

1. Introduction

Projections of wildcat successes and failures in a petroleum play are reasonably based on the assumption that past drilling history influences future drilling outcomes. But how? Very little empirical statistical work that might suggest an answer in available in the published literature. One possible reason is that drilling data, even data as simple as wildcat successes and failures have a spatial dimension that makes systematic analysis complex. Short of empirical studies that provide a guide to the effects of well location and of the history of well successes and failures on the probability that a yet to be drilled wildcat well will be a success, procedures for projecting returns to exploratory well drilling effort must be based on ad hoc assumptions. Probabilistic models used to forecast undiscovered oil and gas in petroleum plays typically incorporate the assumption that wildcat well outcomes are either mutually independent or functionally dependent. The first ignores effects of well drilling history on future drilling outcomes and the second may be unrealistic.

The point process model of wildcat well drilling proposed here is designed to capture these effects. It incorporates spatial interdependencies of well outcomes but differs in some respects from standard marked spatial point process models. [See Ripley

* Massachusetts Institute of Technology, Cambridge, MA 02139
** Institute of Sedimentary and Petroleum Geology, Geological Survey of Canada, 3303 33rd Street NW, Calgary, Canada T26 2A7

(1986) and (1988) for examples]. While marked point process models usually begin with assumptions about the random nature of locations of points in a plane (or in higher dimensions), we shall assume that well locations are non-random covariates that appear as part of the observed history of the drilling process. In fact they are not random as companies do not drill wildcats randomly. Our focus is on the probability of success of the (n+1)st wildcat in a prespecified location conditional on being given the history of the drilling process for the first n wildcats. The outcome of drilling the (n+1)st wildcat may be influenced by both locations and outcomes of that set of the first n wildcats within a distance window d of the location of the (n+1)st. Thus the observed well outcome history has a temporal dimension.

The interpretation of well locations as an auxiliary statistic–exogenous covariates not generated by a spatial random process–coupled with the fact that wildcats are ordered in time leads to binary logit models that are vastly simpler to analyze than models in which well locations are assumed to be generated by a probabilistic spatial point process. For a comparison of difficulties, examine the study by Stoyan, Kendall and Mecke (Stoyan et al. [1975] Chapter 5) of the spatial pattern of 31 sink holes caused by sulphide Karst processes near the Harz region. Treating well locations as non-random exogenous covariates has another important virtue: the usual spatial point process edge effect problem disappears.

A class of binary logit models that incorporate the effects of outcomes of previously drilled wildcats within a prespecified distance window of the "next" well on that well's outcome is specified in Section 3 and then applied to wildcat data from two petroleum plays located in the Western Canada Sedimentary Basin, the Leduc Reef-Windfall Play and the Swan Hills - South Kaybob Play. Even though the analysis done here is conditioned on treating well locations as non-random covariates, a descriptive study of the spatial point patterns formed by wildcat locations is informative. This is done in Section 4 as a prelude to application of binary logit analysis to the data in Section 5. The aim of this preliminary data analysis is to see if wildcats within each of these plays are clustered, randomly dispersed or more dispersed than random and to discern possible differences in the spatial patterns between dry and successful wildcats. Quadrant counts and Ripley's L(d) function (Ripley [1988]) are employed. A cursory visual examination of the data suggests that wildcats are not drilled in a spatially random pattern over the play area. Statistical analysis confirms this. However, what appears obvious for this particular data set may not be obvious elsewhere. (See Ripley [1988] for a discussion of this issue).

Petroleum exploration prospect analysis is a routine exercise designed to evaluate prospect risk–the probability that a prospect is an economically viable deposit–and to appraise the size of the prospect. The method presented here provides an estimate of the probability that a wildcat drilled to confirm a prospect at a particular location discovers a deposit. Because prospect success probabilities are used to establish drilling priorities, the procedure employed to determine these probabilities directly influences economic returns to exploration programs. As a consequence, it is important to understand both strengths and weaknesses of methods currently available for evaluating prospect risk as a guide to their use and in addition, to devise new methods that compensate for their weaknesses.

There are two distinct schools for prospect risk evaluation, Bayesian and Frequentist. Those who adhere to the Bayesian or Subjectivist school assign personal probabilities to risk factors based on interpretation of available geological and geophysical evidence. Ideally a post-drilling evaluation is carried out in order to determine how well a priori judgements about uncertain risk factors match the outcome of drilling.

The Frequentist approach is principally based on observable data. However, in practice those who adopt it also often use subjective interpretation of geological and geophysical evidence to modify empirical estimates of risk factors. Frequentist methods for appraisal of wildcat risk factors currently in use may be roughly classified by the level and type of data required. A common procedure is to compute an estimate of the overall success rate experienced for a play: the ratio of the number of successful wildcats to the total number of wildcats drilled. While simple to execute and useful as a rough guide to assignment of a success probability, this estimate fails to incorporate information specific to locations of yet to be drilled prospects. In practice, it is often supplemented with information provided by geologists who have experience with prospects similar to that about to be drilled. This information is used to modify subjectively the observed ratio of successes to total wildcats. A more recently proposed procedure is based entirely on exploration data. Each exploratory well is examined to determine why it was either dry or a discovery. Statistics describing geological attributes such as absence of trap, porosity, cap rock or source rock for a play history are considered. Examples of this last approach are found in Lee, Qin and Shi (1989). A strictly Frequentist approach to risk assessment has advantages and disadvantages. While estimates of risk factors based solely on observable data are not subject to personal bias, currently available methods of this type are not geographically specific. It is reasonable to expect that risk factors vary geographically over a play area.

Hohn (1988) applies indicator kriging to data from Kumar's (1985) study of wells in the northwest shelf of New Mexico's Delaware Basin in an interesting exercise that yields iso-contours of success probabilities. There is, however, no explicit temporal ordering of wells in his adaptation of kriging to well successes and dry holes.

The method presented here is specifically designed to provide Frequentist type estimates of wildcat well success probabilities as a function of well location and of exploration history within a spatial window about the location. The following premises guide construction of the model presented in Section 3. Exploratory well risk factors and wildcat success in particular may be dependent on:

(1) wildcat location,

(2) the number of dry and of successful wildcats within a spatial window about a drillable prospect, and

(3) the distance between dry and successful wildcats and the prospect to be drilled.

The ultimate aim of analysis of spatial point patterns of wildcat outcomes is to provide explorationists with a tool for prediction of the probability of success of a wildcat to be drilled at a given location as a function of a play's drilling history. To this end iso-contour plots of probability of success as a function of well location and drilling history for two Canadian plays are presented in Section 5. These plots may be used to provide insight complementary to traditional modes of geological and geophysical analysis of where to drill a wildcat. Iso-contour plots of wildcat success probabilities as a function of both location and well outcome history coupled with measures of sampling error allow identification of future exploration fairways (See Section 6) and may be used to cross-validate results provided by other methods of appraising risk factors.

It is possible to expand the set of explanatory variables used to predict success probability beyond just location effect and well outcome history within a spatial window. However, we restrict this particular study of these two simple sets of variables in order to appraise their effectiveness. We conjecture that including geological variables appropriate to the particular play under study will enhance our ability to provide more precise estimates of wildcat success probabilities. This will be a subject of a future paper.

The definition of the population or play being sampled is critical, because different plays exhibit different spatial patterns of risk. Here two Devonian gas plays from the Western Canada Sedimentary Basin (Reinson et al., 1991) are used to illustrate the application of the risk evaluation procedure proposed here.

The Leduc Reef Complex - Windfall Play consists of the Leduc Formations. The Leduc formation is a biohermal carbonate. The average thickness is about 150 m. The basinal shales and limestones act as the lateral and top seals. The play areal extent is shown in Figure 4.1. The data set consists of 297 wildcats and 58 discoveries.

The Swan Hills Shelf Margin - Kaybob South Play is defined to include all gas pools in stratigraphic traps within the carbonate shelf and platform. The thickness of the formation ranges from 30 to 125 m. The northeast margin dolomitized shelf and reefal limestones form the reservoir, and the overlying Waterways Formation seals the reservoir laterally and vertically. The play definition and its areal extent are shown in Figure 4.2. The data set consists of 414 wildcats and 50 discoveries.

A wildcat is defined as an exploratory well that penetrates the lithological zone defining the play under study. A discovery is defined as either a commerical discovery or a recovery from a drill-stem test.

A principal message of this statistical study of wildcat drilling patterns is that for the data studied here there appears to be spatial dependencies among wildcat well outcomes, both as a function of location and of observed drilling history within a spatial window.

2. THE DATA GENERATING PROCESS

The wildcatting process is analyzed as dependent on observed history in the following way: wildcats are labelled $1, 2, \ldots$ in the order drilled. Associated with each wildcat is a description of the *state* of that well: its location and whether or not it is a discovery or dry.

Wildcat states are defined as follows: let $\underline{x}_i$ denote the coordinates of location of wildcat i and define $y_i = 1$ if wildcat i is a discovery and $y_i = 0$ otherwise. Then

$$\underline{s}_i = (\underline{x}_i, y_i) \tag{1.1}$$

is the state description for wildcat i and after drilling n wildcats, the observed history is

$$\underline{s}^{(n)} = (\underline{s}_1, \ldots \underline{s}_n). \tag{1.2}$$

We shall elliptically use the symbol H_n to denote $\underline{s}^{(n)}$ and distinguish a random variable Y_i from a value y_i assumed by it with a capital letter.

Successes and failures $Y_1, \ldots, Y_n, \ldots$ are made dependent on past history in a fashion to be described shortly. As indicated in section 1, well locations are assumed *not subject to uncertainty.*

The data generating process model is of this form: define

$$p(\underline{x}_i \mid H_{i-1}) = \text{ Prob } \{Y_i = 1 \mid \underline{x}_i; H_{i-1}\}. \tag{1.3}$$

That is, given a history H_{i-1} and a wildcat to be drilled at location $\underline{x}_i$, the probability that this well is a success is $p(\underline{x}_i | H_{i-1})$. The joint probability of realizing $(Y_1, \ldots, Y_n) =$

$(y_1, \ldots, y_n)$ is representable as

$$\prod_{i=1}^{n} \left[p(\underline{x}_i \mid H_{i-1})\right]^{y_i} \left[1 - p(\underline{x}_i \mid H_{i-1})\right]^{1-y_i}. \tag{1.4}$$

The probability law represented by (1.4) is flexible enough to incorporate several interesting types of dependencies:

(1) dependence of a wildcat outcome on the *location* of the well

(2) dependence of the $(n+1)$st wildcat outcome on *distances* of wildcats $1, 2, \ldots, n$ from the $(n+1)$st well

(3) dependence of the outcome of the $(n+1)$st wildcat on outcomes of wildcats $1, 2, \ldots, n$.

Dependence on location may be a geological necessity when, for example, drilling history shows a high success ratio on an anticlinal trend and a low success ratio off trend. The model may be specified so that the probabilities of a drilling success depends on well locations but is independent of outcomes of earlier wells. Such a model represents a Bernoulli process with varying probabilities and constitutes a particular form of trend surface analysis. This particular model may be taken as a descriptive null hypothesis against which we wish to test the alternative that the $(n+1)$st wildcat outcome depends on both the locations *and* outcomes of wildcats $1, 2, \ldots, n$ within a spatial window around the $(n+1)st$ wildcat.

3. SPECIFIC MODELS

In order to estimate model parameters and to project future values of success probabilities from knowledge of a well history H_n, $p(\underline{x}_i \mid H_{i-1})$ must be made a specific function of $\underline{x}_i$ and H_{i-1}. Parsimony in choice of the number of parameters is highly desirable. To this end we employ only simple functions to represent dependence on location and on past well outcomes. The generic form for $p(\underline{x}_i \mid H_{i-1})$ we shall study here is

$$p(\underline{x}_i \mid H_{i-1}) = \frac{\exp\{h(\underline{x}_i) + g(\underline{x}_1, \ldots, \underline{x}_i; y_1, \ldots, y_{i-1})\}}{1 + \exp\{h(\underline{x}_i) + g(\underline{x}_1, \ldots, \underline{x}_i; y_1, \ldots, y_{i-1})\}}; \tag{3.1}$$

this is a particular case of a *logit model* in which $h(\underline{x}_i)$ represents the effect of the location of the i^{th} well on its probability of success and $g(\underline{x}_1, \ldots, \underline{x}_i; y_1, \ldots, y_{i-1})$ represents the effect on this probability of interactions between the location of the i^{th} well and locations and outcomes of wells $1, 2, \ldots, i-1$.

A good choice of a particular form for $h(\underline{x}_i)$ depends very much on the geological setting. The choice of $g(\cdot, \cdot)$ is more delicate. This latter function may be chosen so that the outcome of wildcat i depends in some fashion on:

(1) outcomes of wildcats $j = 1, 2, \ldots, i-1$ alone independent of the locations of these wells ,

(2) locations *and* outcomes of all wildcats $j = 1, 2, \ldots, i-1$,

(3) outcomes only of wildcats $j = 1, 2, \ldots, i-1$ that are within a distance d of the location of the i^{th} wildcat,

(4) locations and outcomes within a distance d of the location of the i^{th} wildcat.

The intuition behind choice of a distance window d is that the influence of outcomes of wells further away than d from a well to be drilled is negligible.

In order to simplify exposition momentarily assume that no trend effect is present $(h(\underline{x}_i) \equiv 0)$. Also assume that only *paired* distances between wildcats i and $j = 1, 2, \ldots, i-1$ influence the probability of sucess of the ith wildcat. Defining $d(i,j)$ as the distance of wildcat i from wildcat j suppose that the function g can be expressed in the form

$$g(\underline{x}_1, \ldots, \underline{x}_i; y_1, \ldots, y_{i-1}) = \sum_{j=1}^{i-1} \varphi(d(i,j), y_j). \tag{3.2}$$

The function φ incorporates the effect of both the distance of wildcat j from wildcat i and the outcome of wildcat j on the probability that wildcat i is successful.

A particularly simple choice for φ is this: let

$$\delta(j) = \begin{cases} a & \text{if } Y_j = 1 \\ b & \text{if } Y_j = 0 \end{cases}, \tag{3.3}$$

and

$$\xi(i,j;d) = \begin{cases} 1 & \text{if } d(i,j) \le d \\ 0 & \text{otherwise} \end{cases}. \tag{3.4}$$

Choice of $a \ne b$ implies that the impact of a successful earlier wildcat on the probability that the i^{th} wildcat is successful is different from the impact of a dry hole. Among the first $i-1$ wells the number of wells for which $d(i,j) \le d$ is

$$\sum_{j=1}^{i-1} \xi(i,j;d) \equiv n_{i-1}(d). \tag{3.5}$$

Setting

$$\varphi(d(i,j), y_j) = \xi(i,j;d)\delta(j) \tag{3.6}$$

the number of successful wells for which $D(i,j) \leq d$ is

$$\frac{1}{a} \sum_{j \in \{k \mid \delta_k = a, 1 \leq k \leq i-1\}} \xi(i,j;d)\delta(j) \equiv r_{i-1}(d). \tag{3.7}$$

With $c = a - b$,

$$\sum_{j=1}^{i-1} \varphi(d(i,j), y_j) = ar_{i-1}(d) + b[n_{i-1}(d) - r_{i-1}(d)] = bn_{i-1}(d) + cr_{i-1}(d) \tag{3.8}$$

so that

$$p(\underline{x}_i \mid H_{i-1}) = \frac{e^{bn_{i-1}(d)+cr_{i-1}(d)}}{1 + e^{bn_{i-1}(d)+cr_{i-1}(d)}} \tag{3.9}$$

and with $s_n \equiv \{i \mid y_i = 1 \text{ for } i = 1, 2, \ldots, n\}$ (1.4) is representable as

$$\mathcal{L}(b, c \mid d; H_n) = \prod_{i \epsilon s_n} e^{bn_{i-1}(d)+cr_{i-1}(d)} \prod_{i=1}^{n} \left(1 + e^{bn_{i-1}(d)+cr_{i-1}(d)}\right)^{-1}. \tag{3.10}$$

Interpreted as a function of parameters b and c, $\mathcal{L}$ is the likelihood function for a logit model. Estimators of b and c derived from $\mathcal{L}$ depends on choice of the distance window d via the statistics $r_{i-1}(d)$ the number of successful wildcats among wildcats $1, 2, \ldots i - 1$ that are within distance d of the location of the i^{th} wildcat and the total number $n_{i-1}(d)$ of wildcats within distance d of the i^{th} wildcat among wildcats $1, 2, \ldots, j - 1$.

The behavior of (3.9) as a function of the number of dry holes $n_{i-1}(d) - r_{i-1}(d)$ and the number $r_{i-1}(d)$ of successes within d kilometers of well i depends on magnitudes and signs of a and b. If, for example, $a > 0$ and $b < 0$, then ceteris paribus, an increase in $r_{i-1}(d)$ increases log odds $(\underline{x}_i) = \log[p(\underline{x}_i|H_{i-1})/(1 - p(\underline{x}_i|H_{i-1})]$ and an increase in

$n_{i-1}(d) - r_{i-1}(d)$ decreases log odds $(\underline{x})_i$. If both $a, b > 0$ then log odds$(\underline{x}_i)$ increases with an increase in the number of wells within d kilometers of well i, but at possibly different rates for successes and dry holes – loosely interpretable as a geological learning effect. This is the case for Swan Hills. Leduc is peculiar: $a < 0$ and $b > 0$ so that $c \leq 0$.

The particular choice (3.8) for φ incorporates the effect of wildcat outcomes $y_1, \ldots, y_{i-1}$ within distance d on the probability that wildcat i will be successful, but does not weight outcomes within the distance window d by their distances from wildcat i. To incorporate this type of distance effect consider

$$\varphi(d(i,j), y_j) = \xi(i,j;d)\delta(j)/d(i,j). \tag{3.11}$$

Then

$$\sum_{j=1}^{i-1} \varphi(d(i,j), y_j) = \sum_{j=1}^{i-1} \xi(i,j;d)\delta(j)/d(i,j). \tag{3.12}$$

In analogy to the definitions of $r_i(d)$ and $n_i(d)$ define

$$\rho_{i-1}(d) = \frac{1}{a} \sum_{j \in \{k | \delta_k = a, 1 \leq k \leq i-1\}} \xi(i,j;d)\delta(j)/d(i,j), \tag{3.13}$$

and

$$\eta_{i-1}(d) = \sum_{j=1}^{i-1} \xi(i,j;d)/d(i,j) \tag{3.14}$$

so that (3.12) can be written as

$$\sum_{j=1}^{i-1} \varphi(d(i,j), y_j) = c\rho_{i-1}(d) + b\eta_{i-1}(d). \tag{3.15}$$

In order to compute well outcome-distance effects captured by $bn_{i-1}(d) + cr_{i-1}(d)$ or by $b\eta_{i-1}(d) + c\rho_{i-1}(d)$ a record of inter-well distances is required:

$$
\begin{array}{c|c}
 & \begin{array}{ccccc} 1 & 2 & 3 & \ldots & n \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ \vdots \\ n-1 \\ n \end{array} &
\begin{pmatrix}
0 & d(1,2) & d(1,3) & \ldots & d(1,n) \\
 & 0 & d(2,3) & \ldots & d(2,n) \\
 & & 0 & \ldots & d(3,n) \\
 & & & \ddots & \vdots \\
 & & & 0 & d(n-1,n) \\
 & & & & 0
\end{pmatrix}
\end{array}
$$

Armed with well locations, well outcomes and the above inter-well distance matrix computation of MLE's of parameters b, c and of a particular specification of $h(\underline{x})$ for well location effect is straightforward.

Specifiation of $h(\underline{x})$ should be tailored to fit general features of the spatial pattern of success rates that are apparent in the data. An initial view of the spatial pattern of success rates in 50 $\times$ 50 kilometer quadrants suggest that $h(\underline{x})$ for Leduc should be "hill-like" in Northeastern quadrants and should be "trough-like" for Swan Hills in Northeastern quadrants [See Tables 4.1c, 4.2c and Table 4.3].

In order to capture trough-like behavior of success rates along the diagonal $x = y$, the function $h(\underline{x})$ can be chosen so that $h(\underline{x}) \longrightarrow -M, M >> 0$, as the x co-ordinate approaches the y-coordinate. With $\underline{x} = (x,y)$, $a < 0$ and $\mid a \mid /c = M$, define

$$h((x,y)) = \frac{a}{\mid x - y \mid + c} \tag{3.16}$$

so that $h((x,x)) = -M$. For M very large $\exp\{h(x,x)\}/[1 + \exp\{h((x,x))]\}$ is close to zero. This behavior of h is maintained if a polynomial function of (x,y) co-ordinates is added to (3.16). Upon specifying an a priori value for c and defining $z(x, y; c) = (\mid x - y \mid + c)^{-1}$, z can be treated as an independent explanatory variable in a standard linear logit formulation of $p(\underline{x}_i \mid H_{i-1})$. For example,

$$\log\left[\frac{p(\underline{x}_i \mid H_{i-1})}{1 - p(\underline{x}_i \mid H_{i-1})}\right] = P_m(x, y) + \gamma z(x, y; c) \tag{3.17}$$

where $P_m(x, y)$ is a polynomial of degree m in x and y and γ is a parameter.

4. Description of the Data

Wildcat well data for the Leduc and Swan Hills plays are used to illustrate how the model described in Section 3 can be employed. A statistical description of the data is done as a prelude. Even though the model outlined in Section 3 treats wildcat locations as non-random covariates, an examination of their spatial pattern is informative. To this end temporarily assume that wildcat locations are generated by a spatially random process. It is then appropriate to ask:

(1) Are well locations "random" or clustered?

(2) How does the intensity of wildcats drilled vary as a function of location?

(3) What are answers to (1) and (2) for successful wildcats?

The methods employed to answer these questions are simple: first, quadrant counts and second, a method proposed by Ripley [(1981), (1988)]. Ripley suggests a plot of empirical estimates of the function $L(d)$ vs d as shown in Figures 4a and 4b. The function $L(d)$ is proportional to the square root of the expectation of the number of wildcats within distance d of a well location at $x - y$ co-ordinates when wildcat locations are assumed to have been generated by a Poisson process with constant intensity over the area of the play. An estimate $\hat{L}(d)$ of $L(d)$ may be computed by counting the observed number of wildcats within distance d of each wildcat location. If a plot of $\hat{L}(d)$ vs. d appears to be a straight line, it is plausible that locations are spatially uniformly distributed. Significant curvature outside of 95% confidence bounds (computed assuming that wildcat locations are uniformly random over the play area) signals a departure from a uniformly random distribution of locations (See Ripley [1981] Section 8.3 for details).

Both methods support the conclusions that neither Leduc nor Swan Hills wildcat well locations are uniformly random over their respective play areas and that wildcats are clustered. Successful wildcat locations behave similarly.

4.1 Quadrant Counts

Figure 4.1 shows 297 Leduc wildcat locations with $x - y$ coordinates expressed in kilometers of distance from an arbitrary origin. Wells appear to increase in density as one moves Northeast. Numbers of wildcats in 50×50 kilometer grid squares are displayed in Table 4.1a and numbers of discoveries in Table 4.1b. While it is possible to test the hypothesis of uniformity (null hypothesis that each grid square has an equal number of wildcats in it), against alternatives, the lack of uniformity of numbers of wildcats/grid square is so evident that such a test is redundant.

[Table 4.1a, b, c here] [Figure 4.1 here]

The ratio of Leduc successful wildcats to total wildcats within each 50×50 kilometer quadrant–the success rate–is shown in Table 4.1c. Successes are concentrated in quadrants on or above Northwest to Southeast main diagonal of the table. Quadrants with y co-ordinates -100 to -50 kilometers and quadrants with x co-ordinates -100 to -50 kilometers show no successes at all. While it is tempting to condemn this acreage based on a zero success rate, only nine wildcats have been drilled in these quadrants. This raises a question about modeling tactics: Should a model of the form proposed in Section 3 be fit to the entire play area or should it be fit to a trimmed down area that

TABLE 4.1

LEDUC

TABLE 4.1a
WILDCATS BY QUADRANT

	<-50	-50 to 0	0 to 50	50 to 100	Total
50 to 100	5	26	24	52	107
0 to 50	1	15	73	69	158
-50 to 0	0	2	26	1	29
<-50	0	0	2	1	3
Total	6	43	125	123	297

TABLE 4.1b
SUCCESSES BY QUADRANT

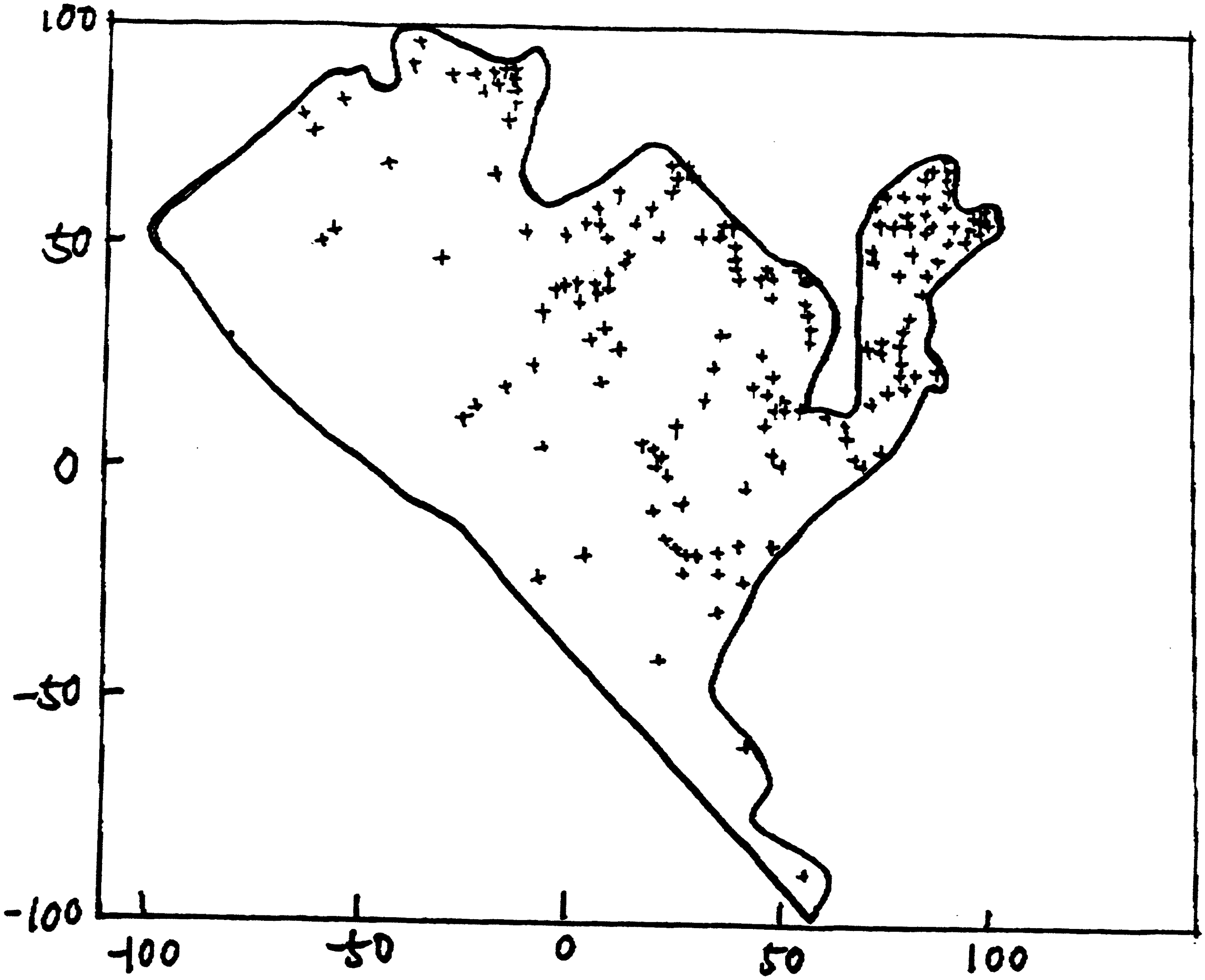
	<-50	-50 to 0	0 to 50	50 to 100	Total
50 to 100	0	5	6	13	24
0 to 50	0	2	20	10	32
-50 to 0	0	0	2	0	2
<-50	0	0	0	0	0
Total	0	7	28	23	58

TABLE 4.1c
SUCCESS RATE BY QUADRANT

	<-50	-50 to 0	0 to 50	50 to 100
50 to 100	0	.1923	.2500	.2500
0 to 50	0	.1333	.2740	.1449
-50 to 0	0	0	.0769	0
<-50	0	0	0	0

FIGURE 4.1

LEDUC REEF COMPLEX - WINDFALL PLAY



[] Success
+ Dry

excludes some acreage to the West and to the South?

Figure 4.2 displays 414 Swan Hills wildcat locations. This figure and the quadrant pattern of Swan Hills wildcats in Tables 4.2a and b share a feature of Leduc wildcats: sparse drilling in far West and far South quadrants and dense drilling in Northeast quadrants. Success rates per quadrant as shown in Table 4.2c, however, exhibit a "trough" on the Southwest to Northeast diagonal, above average success rates off of this diagonal in Northeastern quadrants and low success rates in Western and Southern quadrants. Seven quadrants show a success rate of zero.

[Tables 4.2a, b, c here] [Figure 4.2 here]

Table 4.3 highlights quadrants with above average success rates in both plays. Above average Swan Hills success rates are concentrated in four quadrants off of the Southwest to Northeast diagonal. Leduc above average success rates are concentrated in three Northeastern quadrants. In both plays the maximum positive difference between quadrant success rates and the average success rate is only .08, a small difference that foreshadows difficulty in specifying a model which will provide a sharp increase on predictive accuracy over average success rates.

[Table 4.3 here]

Two features of Figures 4.3a and b, Swan Hills and Leduc discoveries stand out: first, a large fraction of successful wildcats appear in clusters and second, the bulk of discoveries are located in Northeast quadrants above $Y = 0$ and $X = 0$.

[Figure 4.3a and 4.3b here]

TABLE 4.2

SWAN HILLS

TABLE 4.2a
WILDCATS BY QUADRANT

	<-50	-50 to 0	0 to 50	50 to 100	Total
50 to 100	2	28	77	91	198
0 to 50	0	15	31	106	152
-50 to 0	0	1	17	17	35
<-50	0	0	2	27	29
Total	2	44	127	241	414

TABLE 4.2b
SUCCESSES BY QUADRANT

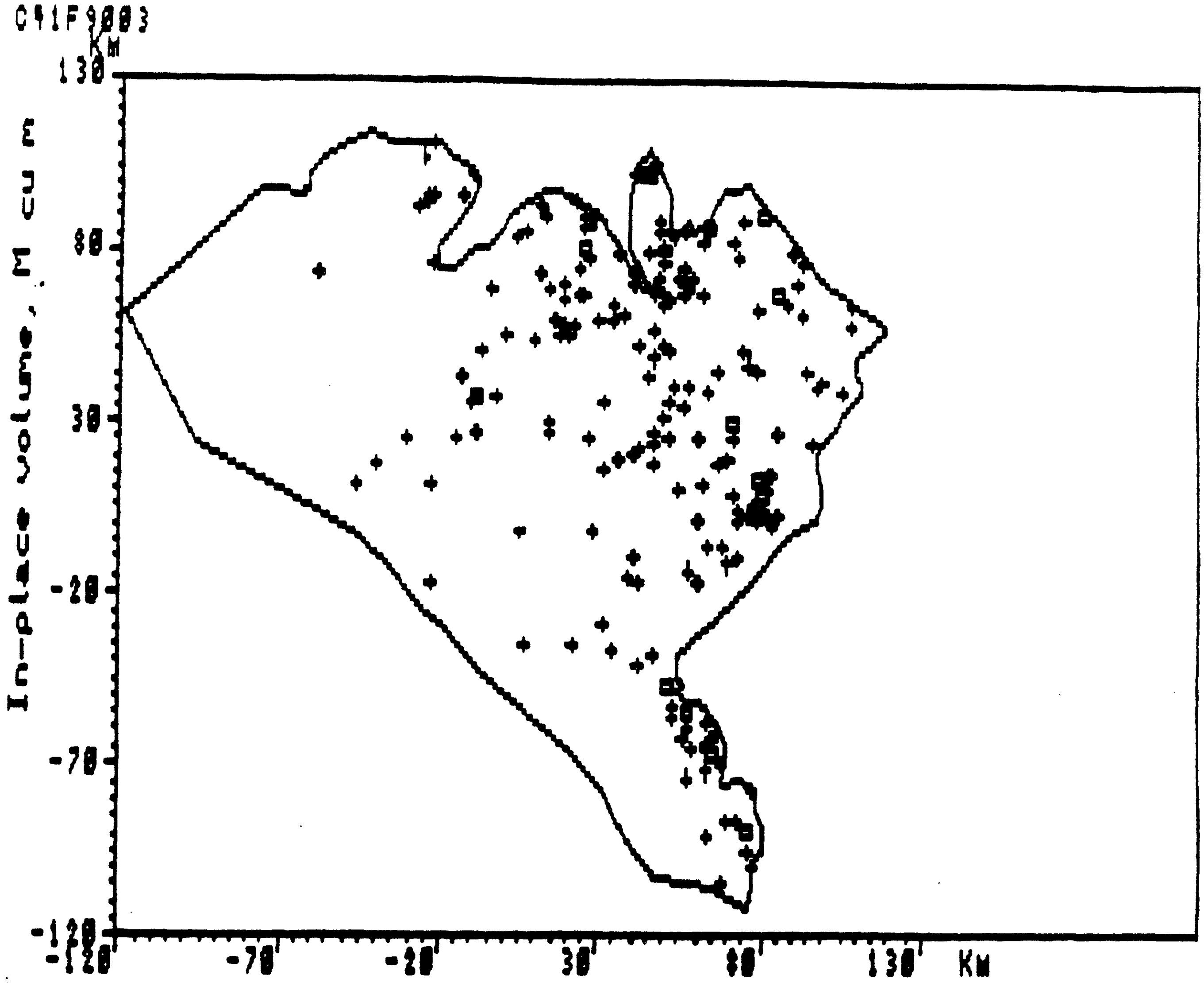
	<-50	-50 to 0	0 to 50	50 to 100	Total
50 to 100	0	2	15	7	24
0 to 50	0	3	2	14	19
-50 to 0	0	0	1	3	4
<-50	0	0	0	3	3
Total	0	5	18	27	50

TABLE 4.2c
SUCCESS RATE BY QUADRANT

50 to 100	0	.0714	.1948	.0769
0 to 50	0	.2000	.0645	.1321
-50 to 0	0	0	.0588	.1765
<-50	0	0	0	.1111

FIGURE 4.2

SWAN HILLS - KAYBOB SOUTH PLAY, DEVONIAN WESTERN CANADA, CANADA



[] Success
+ Dry

TABLE 4.3

CONTRAST = QUADRANT SUCCESS RATE - OVERALL SUCCESS RATE

SWANHILL LAKE

-.12	-.05	+.07	-.04
-.12	+.08	-.06	+.01
-.12	-.12	-.06	+.06
-.12	-.12	-.12	-.01

OVERALL SUCCESS
RATE = .1208

LEDUC +.05

-.2	-.0	+.05	+.05
-.2	-.06	+.08	-.05
-.2	-.2	-.12	-.2
-.2	-.2	-.2	-.2

OVERALL SUCCESS
RATE = .1953

FIGURE 4.3a

SWAN HILLS SUCCESSES

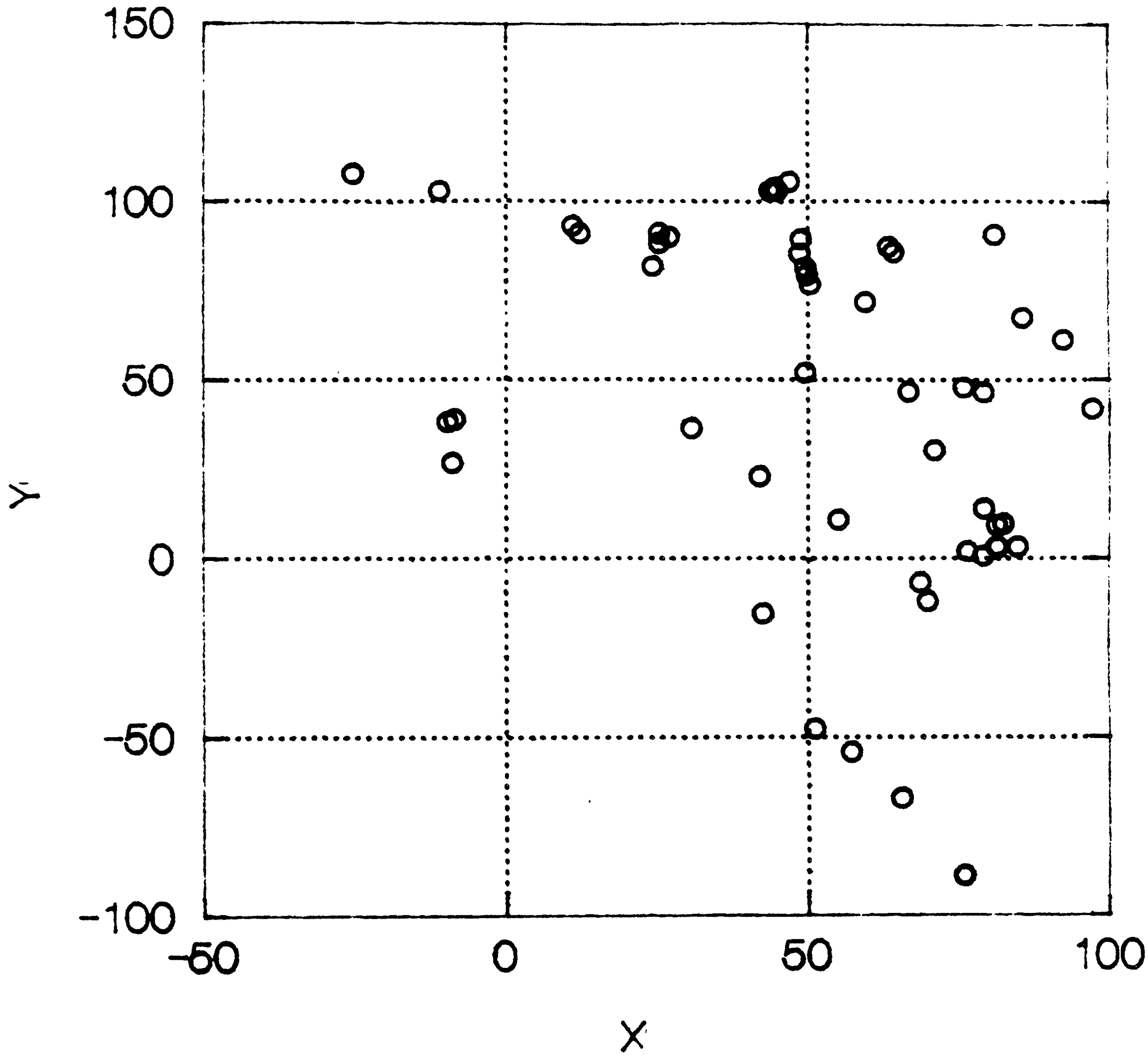
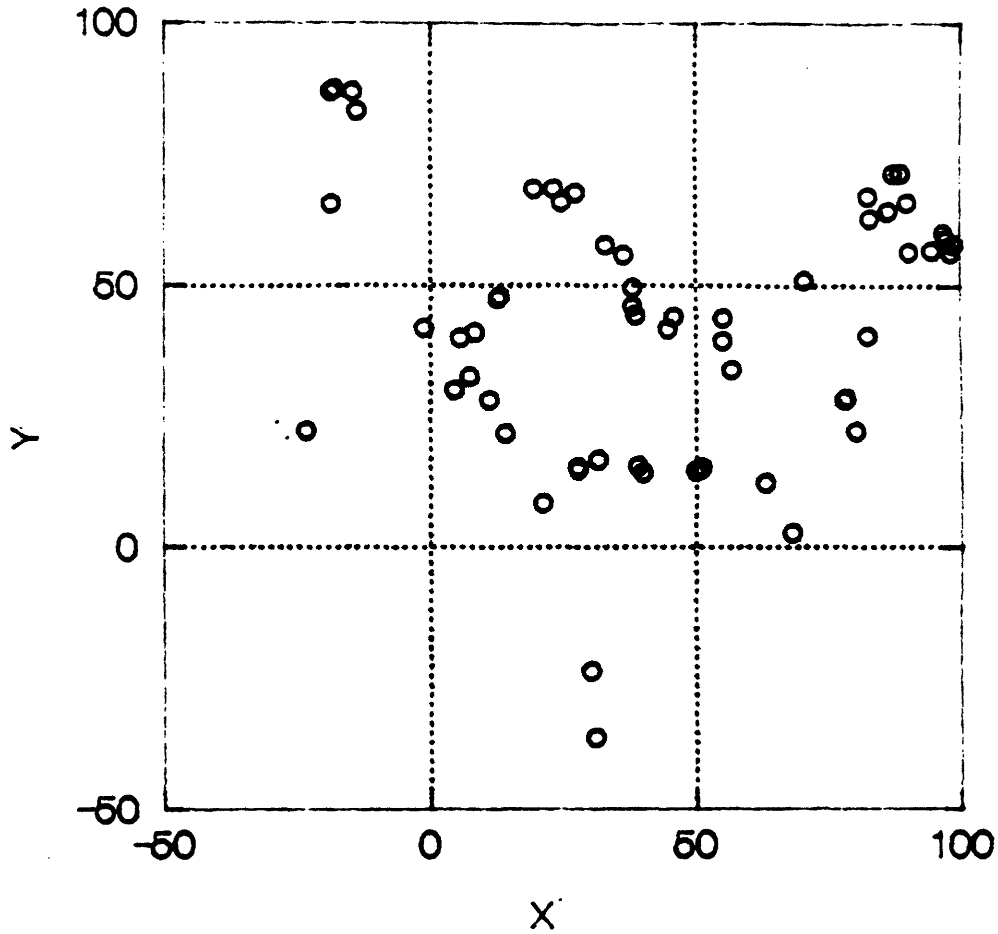


FIGURE 4.3b

LEDUC SUCCESSES



4.2 Dispersion Characteristics

The shape of $\hat{L}(d)$ vs. d for the Leduc play in Figure 4.4 is concave. For inter-well distances less than ten kilometers it lies above the line defining an upper 95% confidence bound for a uniform distribution of well locations. This feature of $\hat{L}(d)$ signals a strong clustering effect. As distance d increases beyond ten kilometers $\hat{L}(d)$ curves outside of the straight line defining a lower 95% bound for a uniform distribution of well locations. A literal interpretation is that for inter-well distances greater than ten kilometers, wells appear to be **more** dispersed than random. This latter feature of $\hat{L}(d)$ is most likely an artifact of **failure to incorporate the impact of edge effects on the behavior of** $\hat{L}(d)$. As the play's boundaries are highly irregular, a proper accounting for edge effects in this data is a substantial computational task. Use of a guard area eliminates too many important well locations and torodial edge correction is unreasonable here. While it is possible to produce an approximately unbiased estimate of the square of $L(d)$ by an inverse weighting scheme (see Ripley [1988] again) our principal focus in this paper is on the behavior of wildcat outcomes when well locations are considered to be non-random exogenous covariates, so we have stopped short of this formidable computation.

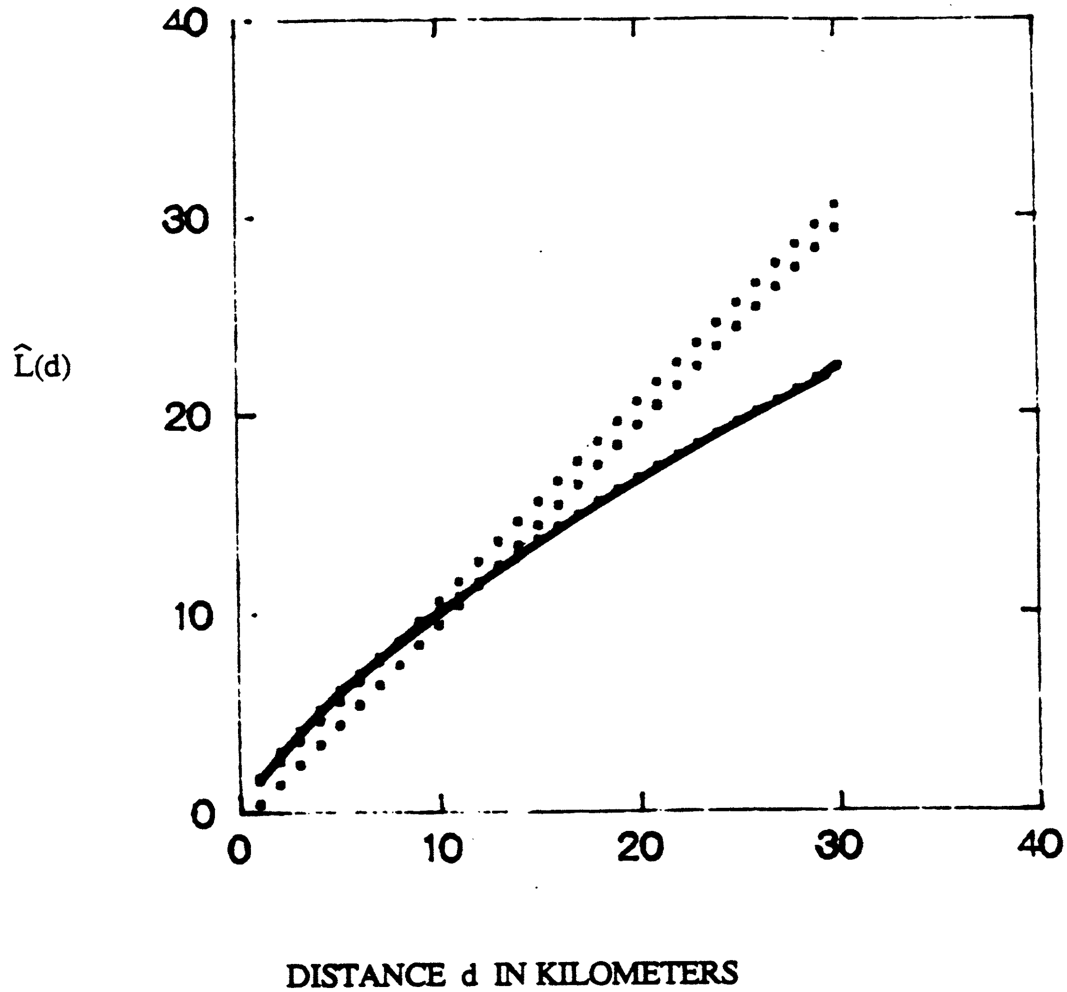
[Figure 4.4 here]

5. Inference Using the Logit Model

Insight into the presence (or absence) of dependence of wildcat well outcomes on past drilling history within a spatial window is afforded by study of the quality of fit of logit models like those discussed in Section 3 to Leduc and Swan Hills data. In keeping

FIGURE 4.4

$\hat{L}(d)$ vs d FOR LEDUC



with simplicity we restrict the location effect function h to be a polynomial function of x- and y-co-ordinates of low degree and adopt one of the two specifications given in Section 3 for interaction of wildcat outcome history with the next wildcat to be drilled. For this data the appropriate choice of statistics is (3.13) and (3.14), distance weighted number of wildcats and distance weighted number of sucessful wildcats within distance d of the next well respectively. Distance weighting of the well history provides a better fit to this particular data.

In addition to investigating spatial dependency it is natural to ask if the model fit improves our ability to predict successs or failure at a particular location relative to a play's overall success rate. Some conclusions about the Leduc and Swan Hills data are:

(1) Use of distance weighted statistics together with the location effect function h specified as a polynomial in x-y co-ordinates provides only a modest increase in playwide prediction of success at actual wildcat locations over the playwide average success rate.

(2) Unweighted numbers of successful and of dry wildcats within a spatial window about the location of the next well to be drilled do not have as strong influence on the probability that this next well will be successful as do distance weighted statistics $\rho_{i-1}(d)$ and $\eta_{i-1}(d)$.

(3) For Swan Hills, $\rho_{i-1}(d)$ and $\eta_{i-1}(d)$ act as surrogates for location effects: incorporating a polynomial location function h adds little explanatory power. For Leduc, however, a better fit to the data is achieved by supplementing

$\rho_{i-1}(d)$ and $\eta_{i-1}(d)$ with a location effect function h.

(4) Iso-contour plots of logit generated estimates of success probabilities as a function of location mirror principal features of successful wildcat locations, suggesting higher than average success probabilities where historical success rates are high and lower than average success rates where historical success rates are low.

Examination of logit regression analyses of Leduc and of Swan Hills data clarifies these points. A simple example is to suppose that the probability of success of the i^{th} wildcat at location $\underline{x}_i$ depends only on the distance weighted statistic $\rho_{i-1}(d)$ as in a model of the form

$$\log \text{ odds } (\underline{x}_i) = \log\left[\frac{p(\underline{x}_i \mid H_{i-1})}{1 - p(\underline{x}_i \mid H_{i-1})}\right] = \text{ constant } + \beta \cdot \rho_{i-1}(d). \tag{5.1}$$

Applied to Swan Hills data with $d = 20$ kilometers and $RECIPS(\underline{x}_i) \equiv \rho_{i-1}(20)$ as defined in (3.13), the estimated model is

$$\log \text{ odds } (\underline{x}_i) = -2.315 + 1.204 * RECIPS(\underline{x}_i). \tag{5.2}$$

According to traditional measures of significance, $RECIPS(\underline{x}_i)$ is a significant explanatory variable, as can be seen from the following summary:

	ESTIMATE	STD. ERROR	T-RATIO	P-VALUE
CONSTANT	-2.315	.190	12.194	.000
RECIPS	1.205	.347	3.474	.001

The odds ratio is proportional to $\exp\{1.204 * RECIPS(\underline{x}_i)\}$ and logit provides confidence bounds for it. On average, a unit change in $RECIPS(\underline{x}_i)$ induces a change of a multiplicative factor $\exp\{1.204\} = 3.333$ in the odds ratio, so

	ODDS RATIO	UPPER 95%	LOWER 95%
RECIPS	3.333	6.574	1.690

Although $RECIPS(\underline{x}_i)$ is clearly an influential explanatory variable, Table 5.1 shows that it provides only a modest 3.1% increase in the average of estimates of success probabilities for wildcats that were in fact discoveries over the entire play's average success rate.

[Table 5.1a, b, c here]

The number 7.588 in the upper left corner of the Table 5.1a in Table 5.1 is computed by summing estimates of probabilities of success $\hat{p}(\underline{x}_i \mid H_{i-1})$ for each of 50 successful Swan Hills wildcats. The number .1518 in the Table 5.1b is the average of this sum (7.588/50 = .1518). In Table 5.1c .1518 is compared to Swan Hills overall average success rate of .1208 (.1518 - .1208 = .0310).

Addition of $RECIP(\underline{x}_i) = \eta_{i-1}(20)$ as an explanatory variable adds virtually no overall explanatory or predictive power as $RECIP$ and $RECIPS$ are highly collinear with sample correlation .581. An artifact of this collinearity is that at $d = 20$ kilometers inclusion of both $RECIP(\underline{x}_i)$ and $RECIPS(\underline{x}_i)$ yields an almost visually indistinguishable iso-contour plot of estimates of success probabilities from that generated by use of $RECIPS(\underline{x}_i)$ alone.

[Figure 5.1 here]

TABLE 5.1

SWAN HILLS

MODEL: LOG ODDS [$p(x_i) \mid H_{i-1}$)] = CONSTANT + RECIPS (x_i), d = 20.

TABLE 5.1a

ACTUAL OUTCOME	PREDICTED OUTCOME: DISCOVERY	PREDICTED OUTCOME: DRY	
DISCOVERY	7.588	42.412	50
DRY	42.412	321.588	364
	50	364	414

TABLE 5.1b

ACTUAL OUTCOME	PREDICTED OUTCOME: DISCOVERY	PREDICTED OUTCOME: DRY	
DISCOVERY	.1518	.8482	1.000
DRY	.1165	.8835	1.000
			1.00

TABLE 5.1c

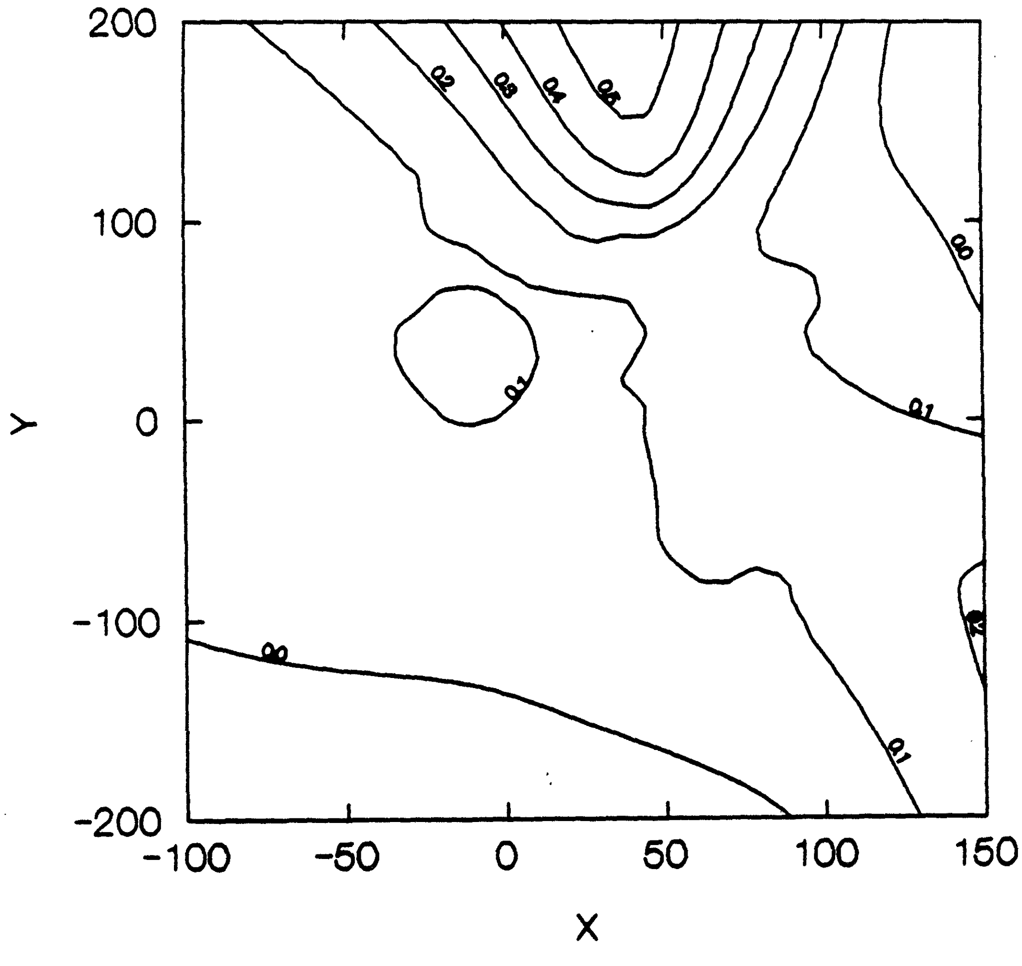
MODEL PREDICTION SUCCESS PROPORTIONS

	CORRECT MODEL PREDICTION	SAMPLE	INCREASE
DISCOVERIES	.1518	.1208	.0310
DRY	.8835	.8792	.0043

PROPORTION OF 414 CORRECTLY PREDICTED = .7951

FIGURE 5.1

SWANHILL d= 20, RECIPS ONLY



Predictive 95% error bounds on estimated probabiliites are provided by SYSTAT's logit regression program. For the Swan Hills data at d =20 kilometers,

	LOWER	$\hat{p}(\underline{x}_i \mid H_{i-1})$	UPPER
MIN	.064	.090	.125
MEAN	.086	.121	.165
MAX	.314	.685	.911

Here is an overall summary of upper and lower error bounds for four wells late in the drilling sequence.

WELL #	LOWER	$\hat{p}(\underline{x}_i \mid H_{i-1})$	UPPER
411	.106	.323	.375
412	.077	.103	.168
413	.069	.096	.179
414	.090	.118	.156

Choice of a large value for d smooths out local variations in success probabilities. The effect of close-by wildcat outcomes on success probability at a given location is captured by reducing the distance window from $d = 20$ kilometers to $d = 5$ kilometers. A fit of $RECIP(\underline{x}_i)$ and $RECIPS(\underline{x}_i)$ to the data for $d = 5$ is

$$\log \text{ odds } (\underline{x}_i) = -2.410 + .501 * RECIP(\underline{x}_i) + .893 * RECIPS(\underline{x}_i) \tag{5.3}$$

with the following standard errors, t-ratios, P-values and odds ratio bounds:

	ESTIMATE	STD. ERROR	T-RATIO	P-VALUE
CONSTANT	-2.413	.212	-11.390	.000
RECIP	.510	.250	2.007	.045
RECIPS	.893	.534	1.674	.094

	ODDS RATIO	UPPER 95%	LOWER 95%
RECIP	1.650	2.692	1.012
RECIPS	2.443	6.952	.859

As for $d = 20$, $RECIP(\underline{x}_i)$ and $RECIPS(\underline{x}_i)$ are collinear with sample correlation .576. The iso-contour plot of success probabilities for $d = 5$ displays a slightly steeper gradient in the Northern-most region of the play and isolates two islands of relatively low success probability [Figure 5.2]. The range of $\hat{p}(\underline{x}_i|H_{i-1})$ across the play is larger for $d = 5$ than for $d = 20$:

	LOWER	PROB	UPPER
MIN	.056	.082	.120
MEAN	.080	.121	.174
MAX	.408	.813	.965

Upper and lower error bounds for the last four wells drilled are:

WELL #	LOWER	$\hat{p}(\underline{s}_i \mid H_{i-1})$	UPPER
411	.263	.551	.808
412	.099	.191	.339
413	.097	.174	.292
414	.073	.106	.152

[Figure 5.2 here]

Addition of a polynomial location function h and/or the function z(x,y;c) defined in (3.16) as a device for incorporating a "trough" in probability iso-contour surfaces gives more flexibility of fit, but provides little additional explanatory power. The message is that for Swan Hills, distance weighted well outcomes statistics are surrogates for location effects.

The Leduc play has quite different statistical properties. For d between 5 and 50 kilometers no adequate fit of a linear combination of $RECIP(\underline{x}_i)$ and $RECIPS(\underline{x}_i)$ alone was obtained. The variables $RECIP(\underline{x}_i)$ and $RECIPS(\underline{x}_i)$ are so highly correlated (.690 for d=5 and .738 for d = 20) as to be surrogates for one another and estimates of the co-efficient of $RECIPS(\underline{x}_i)$ are negative for values of d between 5 and 30, possibly an artifact of collinearity. Iso-contour plots of $\hat{p}(\underline{x}_i|H_{i-1})$ based on these two variables alone show islands of probability that conform roughly to visual clusters of successful wells but do not afford much discrimination. Figure 5.3 is an example for d = 5 kilometers.

[Figure 5.3 here]

FIGURE 5.2

SWANHILL d= 5, RECIP, RECIPS

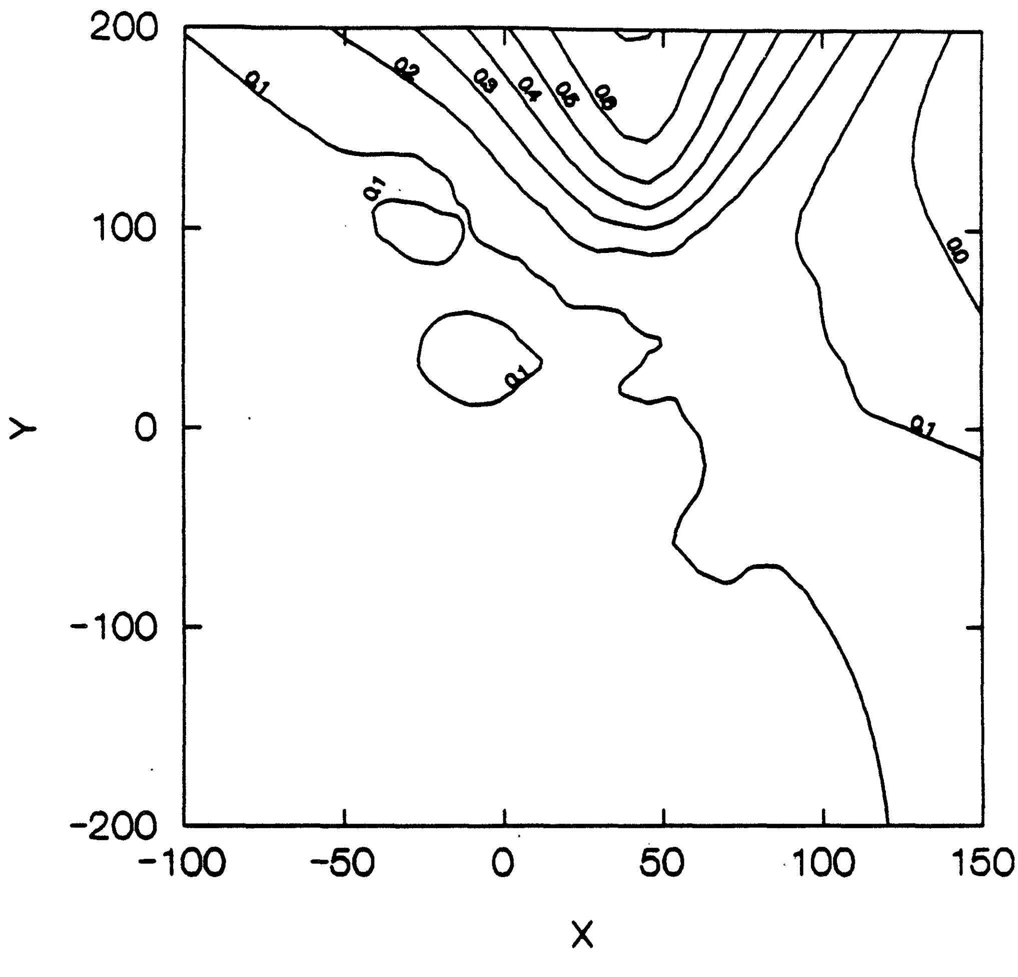
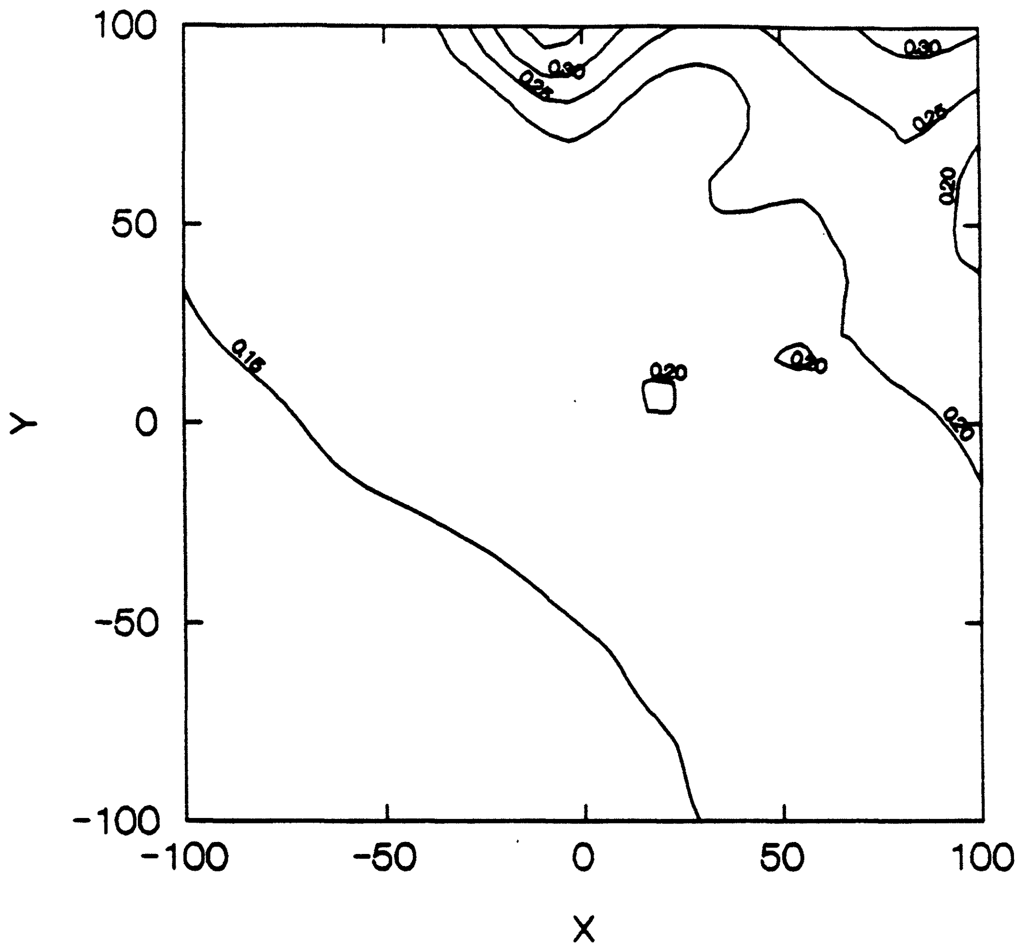


FIGURE 5.3

LEDUC d= 5, RECIP, RECIPS



A fit of $h(\underline{x}) = \text{constant} + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 x^3$ to Leduc data meets standard benchmarks for statistical significance. Estimates of $\beta_1, \ldots, \beta_4$ do not change substantially when $RECIP(\underline{x}_i)$ and $RECIPS(\underline{x}_i)$ are added as explanatory variables:

[Tables 5.2 and 5.3 here]

An iso-contour plot incorporating location effects $h(\underline{x}_i)$ only for Leduc is shown in Figure 5.4 and may be fruitfully compared with Figure 5.5. Coupling $RECIP$ and $RECIPS$ at d = 5 with h results in evident changes in iso-contour patterns in Northeast quadrants of Figure 5.4 as compared to Figure 5.5.

[Figures 5.4 and 5.5 here]

Calculations similar to those in Table 5.1 for Swan Hills show that logit estimation using a location effect function h yields .245 as an average of estimates of probabilities of success for 58 Leduc discoveries wells, and increase of .0436 over playwide success ratio of .2018. If h, $RECIP$ and $RECIPS$ at $d = 5$ are employed, this increase rises to .0592.

Variations in values of d and/or terms included in h generate results that differ in particulars but not in general. As a final example, Figure 5.6 shows iso-contours for $\hat{p}(\underline{x}_i \mid H_{i-1})$ when $d = 20, h(\underline{x}_i) = \text{constant} + \beta_1 x + \beta_2 y + \beta_3 xy + \beta_4 x^2 + \beta_5 y^2 + \beta_6 x^3 + \beta_7 x^3$ and for $d = 20, RECIP(\underline{x}_i)$ and $RECIPS(\underline{x}_i)$ are explanatory variables.

[Figure 5.6 here]

TABLE 5.2

LEDUC-LOCATION EFFECT ONLY

<u>MODEL</u>: log odds $[p(\underline{x}_i|H_{i-1})]$ = constant + x + y $+x^2 + x^3$

	ESTIMATE	STD ERROR	T-RATIO	P-VALUE
CONSTANT	1.624	.419	-3.877	.000
X	.031	.015	2.090	.037
X	.018	.007	2.648	.008
X^2	$-.138 \times 10^{-2}$	$.048 \times 10^{-2}$	-2.893	.004
X^3	$.011 \times 10^{-3}$	$.004 \times 10^{-2}$	2.631	.009

TABLE 5.3

LEDUC-LOCATION EFFECT AND RECIP, RECIPS FOR d=5

<u>MODEL</u>: log odds $[p(\underline{x}_i|H_{i-1})]$ = constant + x + y $+x^2 + x^3$ + RECIP($\underline{x}_i$) + RECIPS($\underline{x}_i$)

	ESTIMATE	STD ERROR	T-RATIO	P-VALUE
CONSTANT	-1.571	.428	-3.668	.000
X	.034	.016	2.120	.034
Y	.017	.007	2.469	.014
X^2	$-.170 \times 10^{-2}$	$.054 \times 10^{-2}$	-3.154	.002
X^3	$.014 \times 10^{-3}$	$.005 \times 10^{-3}$	2.947	.003
RECIP	.447	.229	1.951	.051
RECIPS	-1.126	.554	-2.034	.042

FIGURE 5.4

LEDUC LOCATION EFFECT ONLY

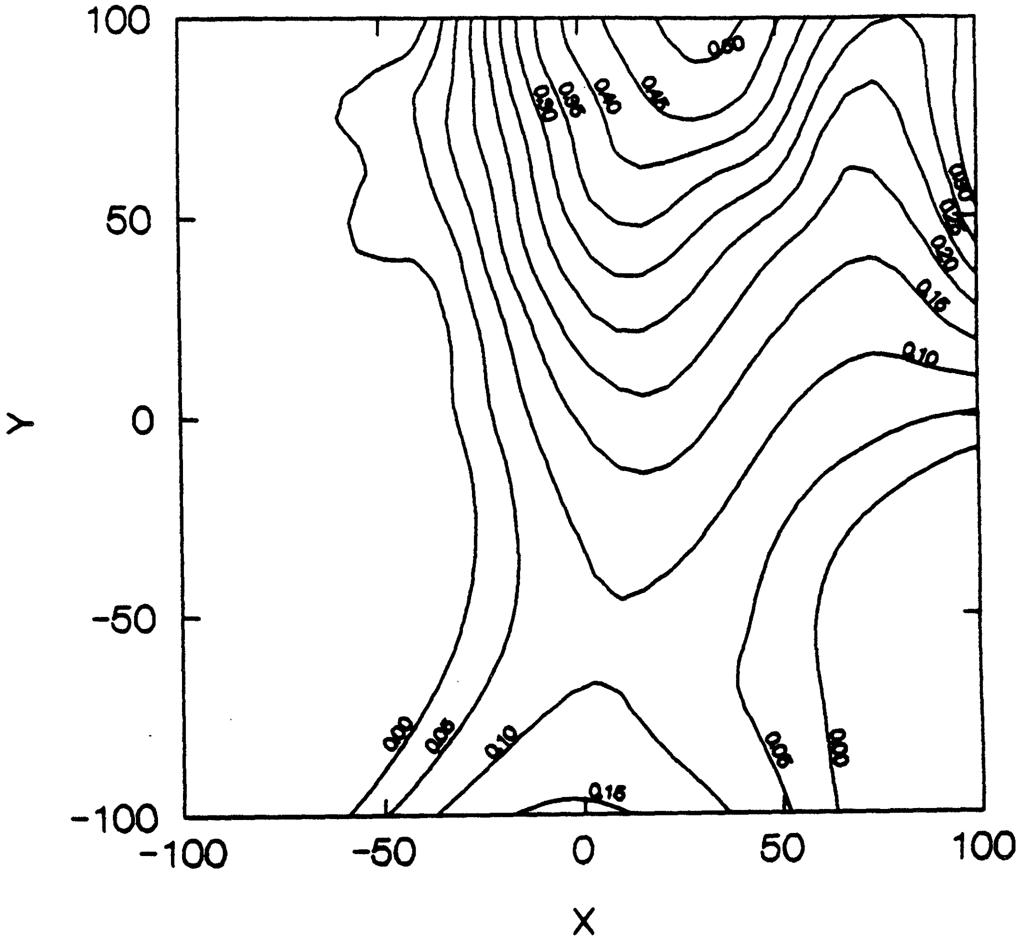


FIGURE 5.5

LEDUC d= 5, RECIP, RECIPS, H(X)

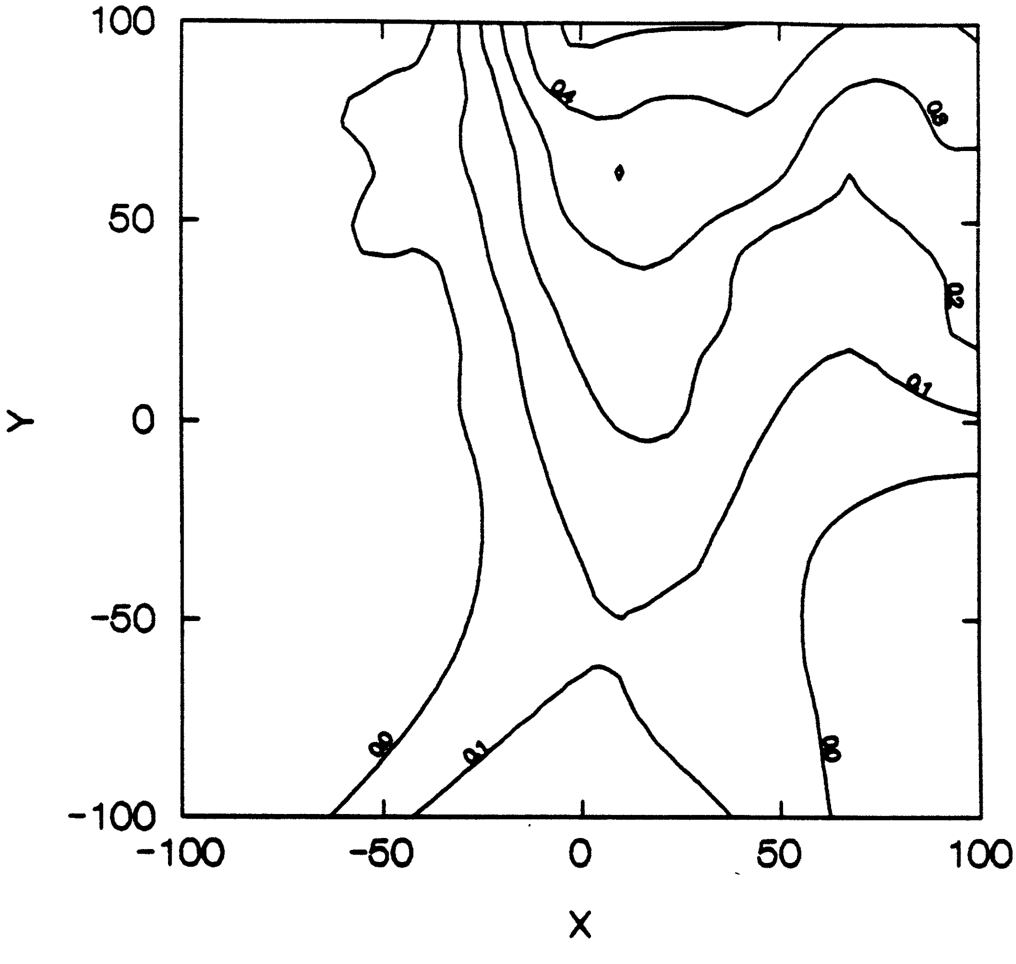
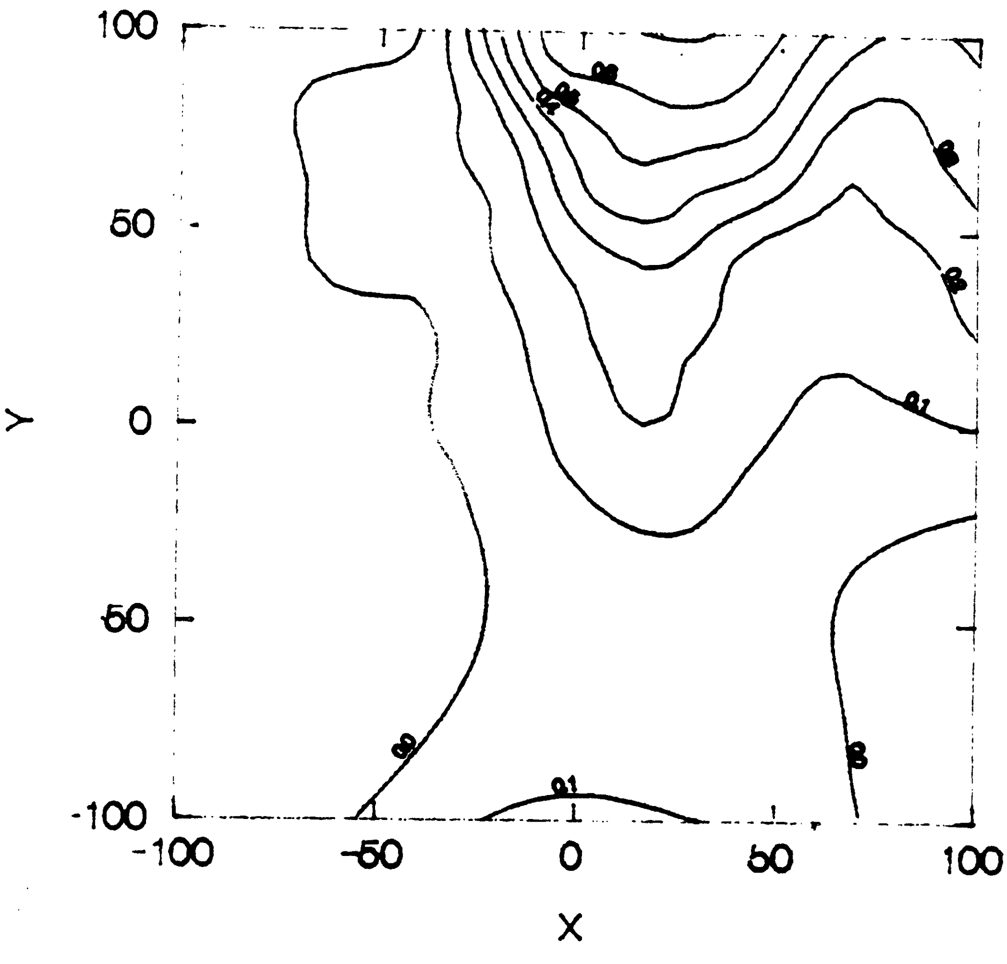


FIGURE 5.6

LEDUC LOGIT PROBABILITIES FOR d=20



References

. M. E., (1988). Geostatistics and Petroleum Geology, Van Nostrand Reinhold, New York.

.umar, N., (1985). Estimating Exploration Potential in Mature Producing Areas, Northwest Shelf of Delaware Basin, New Mexico, *Ball. Amer. Petrol. Geol.*, Vol. 69, p 903-1916.

Lee, P. J., Qin, Ruo-Zhe, and Shi, Yan-Min, (1988). Conditional Probability Analysis of Geological Risk Factors: in Statistical Applications in the Earth Sciences, ed. F. P. Agterberg and G. F. Bonham-Carter; Geological Survey of Canada, Paper 89-9, p. 271-276.

Reinson, G. E., Lee, P. J., Warters, W., Osadetz, K. G., Bell. L. L., Price, P.R., Trollope, F., Campbell, R. I., and Barclay, J.E., (manuscript in preparation). Devonian Gas Resources of Western Canada Sedimentary Basin - Play Definition and Resource Assessment: Geological Survey of Canada, open file.

Ripley, B. D., (1981). Spatial Statistics, John Wiley and Sons, New York.

Ripley, B. D., (1987). Point Process for the Earth Sciences in Quantitative Analysis of Mineral Energy Resources. C., F. Chung et al eds., Reidel Publishing Co.

Stayan D., Kendall, W. S. and Mecke, J., (1987). Stochastic Geometry and its Applications. John Wiley and Sons, New York.